Markov Game with Expected Average Reward Criterion (Markov Game Theory and Their Relative Topics)

AUTHOR(S):

田中, 謙輔

CITATION:

田中, 謙輔. Markov Game with Expected Average Reward Criterion (Markov Game Theory and Their Relative Topics). 数理解析研究所講究録 1982, 460: 229-241

ISSUE DATE:

1982-06

URL:

http://hdl.handle.net/2433/103106

数理解析研究所講究録
460巻 1982年 229-241

Markov game with expected average reward criterion

新潟大・理　田中謙輔

§1. 問題の定式化について

ここでは非協力N人マルコフ・ゲームを（2N+3）個の組$(S, A^1, A^2, \cdots, A^N, p, q, r^1, r^2, \cdots, r^N)$で与えられる：(1) Sは可分な距離空間でゲームの状態空間，(2) A^iはi playerの行動空間と呼ばれるコンパクトな距離空間，(3) pは$S\times A$の上で定義されている正値関数$p(s,\bar{a})$，$\bar{a}=(a^1,a^2,\cdots,a^N)\in\prod_{i=1}^{N}A_i=A$，(4) qは$s\in S$と$\bar{a}\in A$に対応する$(S,\mathcal{B}(S))$の上の確率測度$q(\cdot|s,\bar{a})$，$\mathcal{B}(S)$はSの上の距離によって導入される Borel field である，(5) r^iはi playerの利得率関数と呼ばれる$S\times A$の上で定義されている実数値関数．

このゲームでは各playerが各時点でシステムの状態を観測し，互に相談又は協力することなくて現時点の状態$s\in S$のみで確率的に行動$a^i\in A^i$を選択する．この結果，i player

は利得 $r^i(s,\bar{a})$, $\bar{a}=(a^1,\cdots,a^N)\in A$ を得る．そしてシステムはマルコフ過程に従って新しい状態に移って行く．このとき各 player の最適化問題は 各人の平均期待利得を最大にすることである．

ここでは各 player は現在の状態 $s\in S$ から $P(A^i)$ の中への可測写像である戦略のみを用いることとする，ただし $P(A^i)$ は可測空間 $(A^i, \mathcal{B}(A^i))$ の上のすべての確率測度の集合とする．ゲームシステムの現時点までの歴史に関係していないこのような戦略は定常戦略と呼ばれている．初期状態 $s\in S$ と多重戦略 $\bar{\mu}=(\mu^1,\mu^2,\cdots,\mu^N)$ に対して時刻 T までの player の期待利得は次のように与えられている．

$$\Phi^i(s,T,\bar{\mu})=E_{\bar{\mu}}\left[\int_0^T r^i(s_t,\bar{a}_t)dt \mid s_0=s\right],$$

ただし，s_t, $\bar{a}_t$ は時刻 t におけるシステムの状態と多重行動を示しており，$E_{\bar{\mu}}$ は $\bar{\mu}$ によって決定される確率測度による平均を示している．さらに次のような記号を導入する

$$\phi^i(s,\bar{\mu})=\overline{\lim_{T\to\infty}}\ \frac{\Phi^i(s,T,\bar{\mu})}{T}.$$

このとき，もし

$$\varphi^i(s,\bar{\mu}) = \sup_{\sigma^i} \varphi^i(s,(\mu^{\hat{i}},\sigma^i))$$

となる $\bar{\mu} = (\mu^1, \mu^2, \cdots, \mu^N) \in \prod_{i=1}^{N} P(A^i) = P(A)$ が存在すれば，この $\bar{\mu}$ はマルコフ・ゲームの平衡点と呼ばれ，さらに各要素 μ^i は i player の平衡戦略と呼ばれている，ただし $(\mu^{\hat{i}}, \sigma^i) = (\mu^1, \cdots, \mu^{i-1}, \sigma^i, \mu^{i+1}, \cdots, \mu^N)$.

§2. 準備と記号

$C(A^i)$ を A^i 上の実数値連続関数の全体とする．$F(A^i)$ を $(A^i, \beta(A^i))$ 上のすべての signed measure の集合とし，$C(A^i)$ の双対空間から弱位相を導入する．このとき $F(A^i)$ は linear Hausdorff locally convex topological space となり $P(A^i)$ は $F(A^i)$ の compact convex metric subspace となる．$\prod_{i=1}^{N} F(A^i)$ も linear Hausdorff locally convex topological space となり，$P(A) = \prod_{i=1}^{N} P(A^i)$ は compact convex metric subspace となるので，$P(A)$ は可分になっている．

さらに次のような仮定を与える

(A1)(i) $\varphi(s,\bar{a})$ は各 $\bar{a} \in A$ に対して $\beta(S)$ 可測で，各 $s \in S$ に対して，A 上で連続となっている．さらに次のような正数 M が存在する：

$$0 < p(s,\bar{a}) \leqq M \quad {}^{\forall}s\in S,\ {}^{\forall}\bar{a}\in A .$$

(ii) $q(\cdot\,|\,s,\bar{a})$ は，各 $s\in S$ と $\bar{a}\in A$ に対して，$(S,\mathcal{B}(S))$ 上の確率測度である．各 $\bar{a}\in A$ と $\Lambda\in\mathcal{B}(S)$ に対して，q は $\mathcal{B}(S)$ 可測で，各 $\Lambda\in\mathcal{B}(S)$ と $s\in S$ に対して，A 上で連続となっている．さらに，次のような条件をみたしている：${}^{\forall}s\in S,\bar{a}\in A$, $\Lambda\in\mathcal{B}(S)$,

$$0\leqq q(\Lambda\,|\,s,\bar{a})\leqq 1,$$

$$q(\{s\}\,|\,s,\bar{a}) = 0,$$

$$q(S\,|\,s,\bar{a}) = 1 .$$

(Aｚ) $r^{i}(s,\bar{a})$ は，各 $\bar{a}\in A$ に対して，$\mathcal{B}(S)$ 可測で，各 $s\in S$ に対して，A 上の連続関数となっている．さらに次のような条件をみたす正数 N が存在する：

$$|r^{i}(s,\bar{a})|\leqq N \quad {}^{\forall}s\in S,\ \bar{a}\in A .$$

ここで次のような記号も導入する：$P(A)\ni\mu$ に対して，

$$Q(\Lambda, s, \bar{\mu}) = \int_A p(s,\bar{a})\, q(\Lambda \mid s, \bar{a})\, d\bar{\mu}(\bar{a})$$

$$= \int_{A^1} \cdots \int_{A^N} p(s, a^1, \cdots, a^N)\, q(\Lambda \mid s, a^1, \cdots, a^N) \prod_{i=1}^{N} d\mu^i(a^i),$$

$$r^i(s, \bar{\mu}) = \int_A r^i(s, \bar{a})\, d\bar{\mu}(\bar{a})$$

$$= \int_{A^1} \cdots \int_{A^N} r^i(s, a^1, \cdots, a^N) \prod_{i=1}^{N} d\mu^i(a^i),$$

上の等式は Fubini の定理より結論されている．このとき次の微分方程式は唯一つの解 $p^{\bar{\mu}}(t, \Lambda \mid s)$ a.e. $t \geq 0$ が存在することが知られている： a.e. $t \geq 0$, $\Lambda \in \mathcal{B}(S)$, $s \in S$, $\bar{\mu} \in P(A)$ に対して，

$$\frac{dp^{\bar{\mu}}(t, \Lambda \mid s)}{dt} = -\int_{\Lambda} p(x, \bar{\mu})\, p^{\bar{\mu}}(t, dx \mid s) + \int_S Q(\Lambda, x, \bar{\mu})\, p^{\bar{\mu}}(t, dx \mid s), \tag{2.1}$$

$$p^{\bar{\mu}}(0, \Lambda \mid s) = \delta(s, \Lambda) \qquad \forall \bar{\mu} \in P(A),$$

ただし

$$\delta(s, \Lambda) = \begin{cases} 1 & s \in \Lambda \\ 0 & s \notin \Lambda \end{cases}.$$

さらに，この解 $\phi^{\bar{\mu}}(t, \Lambda | s)$ は次のような性質をみたしている：

(1) $\phi^{\bar{\mu}}$ は各 $t \in [0, \infty)$ と $s \in S$ に対して，$(S, \beta(S))$ 上の確率測度である，

(2) $\phi^{\bar{\mu}}$ は各 $s \in S$ と $\Lambda \in \beta(S)$ に対して，t の絶対連続関数である，

(3) $\phi^{\bar{\mu}}$ は各 $t \in [0, \infty)$ と $\Lambda \in \beta(S)$ に対して $\beta(S)$ 可測になっている．

$M(S)$ は S 上の有界なボレル可測な実数値関数の全体からなる集合とすると，$M(S) \ni u$ に対して $\|u\| = \sup_s |u(s)|$ なる位相の導入によって $M(S)$ は Banach 空間となる．ここで多重戦略 $\bar{\mu} \in P(A)$ に対して，2つの線形写像 $T_t(\bar{\mu})$ と $A(\bar{\mu})$: $M(S) \to M(S)$ を次のように定義する：$u \in M(S)$ に対して

$$T_t(\bar{\mu})u(s) = \int_S u(x)\phi^{\bar{\mu}}(t, dx | s)$$

$$A(\bar{\mu})u(s) = -\phi(s, \bar{\mu})u(s) + \int_S u(x)Q(dx, s, \bar{\mu}).$$

このとき，$T_t(\bar{\mu})$ と $A(\bar{\mu})$ は次の性質をみたしている：

(T1) $\lim_{t\to 0+} T_t(\bar{\mu})u(s) = T_0(\bar{\mu})u(s) = u(s)$

(T2) $T_t(\bar{\mu})T_s(\bar{\mu})u(s) = T_{t+s}(\bar{\mu})u(s)$.

この性質をみたす $\{T_t(\bar{\mu}) : t\in[0,\infty)\}$ は 推移確率 $p^{\bar{\mu}}$ に対応する semi-group と呼ばれている.

Lemma 1. 任意の多重戦略 $\bar{\mu}\in P(A)$ と $u\in M(S)$ に対して

$$\frac{dT_t(\bar{\mu})u(s)}{dt} = T_t(\bar{\mu})A(\bar{\mu})u(s)$$

が成立する.

Proof.

$$\frac{dT_t(\bar{\mu})u(s)}{dt} = \int_S u(x)\frac{dp^{\bar{\mu}}(t,dx|s)}{dt}$$

$$= \int_S u(x)(-p(x,\bar{\mu}))p^{\bar{\mu}}(t,dx|s) +$$

$$+ \int_S u(x)\int_S Q(dx,y,\bar{\mu})p^{\bar{\mu}}(t,dy|s)$$

$$= \int_S \{-p(x,\bar{\mu})u(x) + \int_S u(y)Q(dy,x,\bar{\mu})\}p^{\bar{\mu}}(t,dx|s)$$

$$= T_t(\bar{\mu})A(\bar{\mu})u(s).$$

Lemma 2. 任意の多重戦略 $\bar{\mu} \in P(A)$ と $u \in M(S)$ に対して

$$u(s) = \int_0^\infty e^{-\alpha t} T_t(\bar{\mu})(\alpha I - A(\bar{\mu}))u(s)\,dt$$

が成立する，ただし α は正の実数で，I は恒等写像である．

Proof.

$$\int_0^\infty e^{-\alpha t} T_t(\bar{\mu})(\alpha I - A(\bar{\mu}))u(s)\,dt$$

$$= \int_0^\infty \left\{ \alpha e^{-\alpha t} T_t(\bar{\mu})u(s) - e^{-\alpha t} \frac{dT_t(\bar{\mu})u(s)}{dt} \right\} dt$$

$$= \int_S u(x) \int_0^\infty \frac{d}{dt}\left(e^{-\alpha t} p^{\bar{\mu}}(t, dx \mid s) \right) dt$$

$$= u(s).$$

上のような事実と Ky, Fan の不動点定理より次のような補助定理が成立することが示される．

Lemma 3. $\alpha > 0$ に対して，次の式をみたす多重戦略 $\bar{\mu}_* = (\mu_*^1, \mu_*^2, \cdots, \mu_*^N) \in P(A)$ と $u(\mu_*^{\hat{i}}) \in M(S)$ が存在する：

$$\begin{aligned} \alpha u(\mu_*^{\hat{i}}) &= \max_{\sigma^i} \{ r^i(s, (\mu_*^{\hat{i}}, \sigma^i)) + A(\mu_*^{\hat{i}}, \sigma^i) u(\mu_*^{\hat{i}})(s) \} \\ &= r^i(s, \bar{\mu}_*) + A(\bar{\mu}_*) u(\mu_*^{\hat{i}})(s) \end{aligned} \tag{2.2}$$

ただし $\mu_*^{\hat{i}} = (\mu_*^1, \cdots, \mu_*^{i-1}, \mu_*^{i+1}, \cdots, \mu_*^N) \in \prod_{j \neq i} \mathcal{P}(A^j)$.

§3. 割引因子をもたないマルコフ・ゲームにおける平衡点の存在について

この章では仮説（A1），（A2）に加えて，さらに次のような仮説を課すことにする．

（A3） 次のような条件をみたす正数 γ が存在する：

$$0 < \gamma \leqq \phi(s, \bar{a}) \qquad \forall s \in S, \ \forall \bar{a} \in A .$$

（A4） 次のような条件をみたすある状態 $s_0 \in S$ と正数 $\beta \in (0, 1)$ が存在する：

$$q(\{s_0\} \mid s, \bar{a}) > \beta > 0 \qquad \forall \bar{a} \in A, \ s \neq s_0 .$$

このとき（A3）の条件から (2.1) の解は honest 性をもっている，i.e.,

$$\phi^{\bar{\mu}}(t, S \mid s) = 1 \qquad \forall s \in S, \ \bar{\mu} \in \mathcal{P}(A).$$

今次のような新しい関数 $\bar{\phi}(s, \bar{a})$ と確率測度 $\bar{q}(\cdot \mid s, \bar{a})$ を作る：

$$\bar{\phi}(s, \bar{a}) = \begin{cases} \phi(s, \bar{a}) - \beta\gamma & s \neq s_0 \\ \phi(s, \bar{a}) & s = s_0 , \end{cases}$$

任意の $\Lambda \in \mathcal{B}(S)$ に対して

$$\overline{q}(\Lambda \mid s, \bar{a}) = \begin{cases} q(\Lambda \mid s, \bar{a}) & , \quad s = s_0 \\ \dfrac{p(s,\bar{a}) q(\Lambda \mid s, \bar{a})}{\overline{p}(s, \bar{a})} & , \quad s \neq s_0, \ s_0 \notin \Lambda \\ \dfrac{p(s,\bar{a}) q(\Lambda \mid s, \bar{a}) - \beta\gamma}{\overline{p}(s, \bar{a})} & , \quad s \neq s_0, \ s_0 \in \Lambda . \end{cases}$$

上のような作り方から $\overline{p}(s,\bar{a})$ と $\overline{q}(\cdot \mid s, \bar{a})$ は仮説（A1），（A2）をみたしているので Lemma3 より 次のような多重戦略 $\overline{\mu}_* = (\mu_*^1, \mu_*^2, \cdots, \mu_*^N) \in P(A)$ と $u(\mu_*^{\hat{i}}) \in M(S)$ が存在する： $\beta\gamma > 0$ に対して，

$$\begin{aligned} \beta\gamma u(\mu_*^{\hat{i}})(s) &= \max_{\sigma^i} \{ r^i(s, (\mu_*^{\hat{i}}, \sigma^i)) + \overline{A}(\mu_*^{\hat{i}}, \sigma^i) u(\mu_*^{\hat{i}})(s) \} \\ &= r^i(s, \overline{\mu}_*) + \overline{A}(\overline{\mu}_*) u(\mu_*^{\hat{i}})(s), \end{aligned} \tag{3.1}$$

ただし

$$\overline{A}(\bar{\mu}) u(s) = -\overline{p}(s, \bar{\mu}) u(s) + \int_S u(x) \overline{Q}(dx, s, \bar{\mu}),$$

$$\overline{Q}(\Lambda, s, \bar{\mu}) = \int_A p(s, \bar{a}) \overline{q}(\Lambda \mid s, \bar{a}) d\bar{\mu}(\bar{a}) .$$

このとき (3.1) の中の $\overline{A}$ を A に書き改めることによって (3.1) 式は次のように書き換えられる： $\forall s \in S$ に対して，

$$\beta\gamma u(\mu_*^{\hat{i}})(s_0) = \max_{\sigma^i} \{ r^i(s, (\mu_*^{\hat{i}}, \sigma^i)) + A(\mu_*^{\hat{i}}, \sigma^i) u(\mu_*^{\hat{i}})(s) \} . \tag{3.2}$$

同様にして，次式が得られる

$$\beta\gamma u(\mu_*^{\hat{i}})(s_0) = r^i(s, \overline{\mu}_*) + A(\overline{\mu}_*)u(\mu_*^{\hat{i}})(s). \qquad (3.3)$$

Theorem. 仮説（A1）～（A4）のもとで，割引因子をもたない非協力N人マルコフ・ゲームは平衡点をもっている．

Proof. $g(\mu_*^{\hat{i}}) = \beta\gamma u(\mu_*^{\hat{i}})(s_0)$ とおくと，(3.2) は次のように書ける：$\forall \sigma^i \in P(A^i)$,

$$g(\mu_*^{\hat{i}}) \geq r^i(s, (\mu_*^{\hat{i}}, \sigma^i)) + A(\mu_*^{\hat{i}}, \sigma^i)u(\mu_*^{\hat{i}})(s). \qquad (3.4)$$

(3.4) の両辺に $T_*(\mu_*^{\hat{i}}, \sigma^i)$ を作用し，honest性を用いると

$$g(\mu_*^{\hat{i}}) \geq T_*(\mu_*^{\hat{i}}, \sigma^i)r^i(s, (\mu_*^{\hat{i}}, \sigma^i)) + T_*(\mu_*^{\hat{i}}, \sigma^i)A(\mu_*^{\hat{i}}, \sigma^i)u(\mu_*^{\hat{i}})(s)$$

を得る．

上式にLemma 1を用いて，次に0からTまで積分することから次式に達する：

$$Tg(\mu_*^{\hat{i}}) \geq \int_0^T T_*(\mu_*^{\hat{i}}, \sigma^i)r^i(s, (\mu_*^{\hat{i}}, \sigma^i))dt + T_*(\mu_*^{\hat{i}}, \sigma^i)u(\mu_*^{\hat{i}})(s) - u(\mu_*^{\hat{i}})(s).$$

両辺をTで割って，次に $T \to \infty$ として上極限をとると

$$g(\mu_*^{\hat{i}}) \geq \overline{\lim_{T\to\infty}} \frac{\Phi^i(s, T, (\mu_*^{\hat{i}}, \sigma^i))}{T}$$

$$= \Phi^i(s, (\mu_*^{\hat{i}}, \sigma^i)) \qquad (3.5)$$

が得られる．

(3.3) から同様な議論によって

$$g(\mu_*^{\hat{i}}) = \Phi^{\hat{i}}(s, \overline{\pi}_*) \tag{3.6}$$

が得られる.

(3.5) と (3.6) から

$$\Phi^{\hat{i}}(s, \overline{\pi}_*) = \max_{\sigma^{\hat{i}}} \Phi^{\hat{i}}(s, (\mu_*^{\hat{i}}, \sigma^{\hat{i}}))$$

が成立し, $\overline{\pi}_* \in P(A)$ が マルコフ・ゲームの平衡点になっている. よって証明は完成された.

References

1. Beneš, V.E., Existence of optimal strategies based on specific information for a class of stochastic decision problems, SIAM J. Control and Optimization, 8(1970).
2. Fan,Ky, Fixed points and minimax theorems in locally convex topological linear spaces, Proc. Nat. Acad. Sci., USA, 38(1952).
3. Feller, W., On the integro-differential equations of purely discontinuous Markov processes, Trans. Amer. Math. Soc., 48(1940).
4. Lai, H.C. and Tanaka,K., Non-cooperative n-person game with a stopped set, to appear in J. Math. Analy. Appl.
5. Tanaka,K. and Lai, H.C., Two-person zero-sum game with a stopped set, to appear in J. Math. Analy. Appl.
6. Tanaka,K. and Homma,H., Continuous time non-cooperative n-person Markov games, Bull. Math. Statist., 15(1978).
7. Tanaka,K. and Wakuta,K., On continuous time Markov games with countable state space, J. O.R. Japan, 21(1978).
8. Tanaka,K. and Homma,H., On the learning algorithm of 2-person zero-sum Markov game, Bull. Math. Statist., 19 (1980).
9. Tanaka,K., Markov Game, Proceedings of symposium on functional analysis and applications, at National Tsing Hua University, Taiwan, Republic of China, 1980.
10. Tanaka,K., Stochastic differential games, Report on Research under Financial support of the National Sciences, Council, Republic of China, 1980.